10/531027

PCT/IB03/04522

| Europäisches Patentamt | European Patent Office | Office européen des brevets |
|---|---|---|



| Bescheinigung | Certificate | Attestation |
|---|---|---|
| Die angehefteten Unterlagen stimmen mit der ursprünglich eingereichten Fassung der auf dem nächsten Blatt bezeichneten europäischen Patentanmeldung überein. | The attached documents are exact copies of the European patent application described on the following page, as originally filed. | Les documents fixés à cette attestation sont conformes à la version initialement déposée de la demande de brevet européen spécifiée à la page suivante. |

| Patentanmeldung Nr. | Patent application No. | Demande de brevet n° |
|---|---|---|

02292552.3



Der Präsident des Europäischen Patentamts;
Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets
p.o.

R C van Dijk

Anmeldung Nr:
Application no.: 02292552.3
Demande no:

Anmeldetag:
Date of filing: 16.10.02
Date de dépôt:

Anmelder/Applicant(s)/Demandeur(s):

Koninklijke Philips Electronics N.V.
Groenewoudseweg 1
5621 BA Eindhoven
PAYS-BAS

Bezeichnung der Erfindung/Title of the invention/Titre de l'invention:
(Falls die Bezeichnung der Erfindung nicht angegeben ist, siehe Beschreibung.
If no title is shown please refer to the description.
Si aucun titre n'est indiqué se referer à la description.)

Video encoding method

In Anspruch genommene Prioriät(en) / Priority(ies) claimed /Priorité(s) revendiquée(s)
Staat/Tag/Aktenzeichen/State/Date/File no./Pays/Date/Numéro de dépôt:

Internationale Patentklassifikation/International Patent Classification/
Classification internationale des brevets:

H04N7/24

Am Anmeldetag benannte Vertragstaaten/Contracting states designated at date of filing/Etats contractants désignées lors du dépôt:

AT BE BG CH CY CZ DE DK EE ES FI FR GB GR IE IT LI LU MC NL PT SE SK TR

"VIDEO ENCODING METHOD"

## FIELD OF THE INVENTION

The present invention generally relates to the field of object-based coding, and more particularly to a method for encoding a sequence of video data according to a standard in which several types of data are identified, said data consisting of so-called video Object Planes (VOPs) that are either intra coded VOPs (I-VOPs), coded using information only from themselves, or predictive coded VOPs (P-VOPs), coded using a motion compensated prediction from a past reference VOP, or bidirectionally predicted VOPs (B-VOPs), coded using a motion-compensated prediction from past and future reference VOPs.

## BACKGROUND OF THE INVENTION

The MPEG-4 visual standard provides technologies to view, access and manipulate objects (rather than pixels, with the previous MPEG standards) in a large range of bit rates. The main application areas are for instance : digital television, streaming video, mobile multimedia, games, etc. Said standard operates on video objects (VOs) defined by temporal and spatial information in the form of shape, motion and texture information, coded separately in the bitstream (these VOs are the entities that the user can access and manipulate).

The MPEG-4 approach relies on a content-based visual data representation of the successive scenes of a sequence, each scene being a composition of VOs with its intrinsic properties : shape, motion texture. In addition to the concept of VO, the MPEG-4 standard introduces other ones like the Video Object Layer (each VO can be encoded either in a scalable or non-scalable form, depending on the application, represented by the video object layer, or VOL) and the Video Object Planes (VOPs) (= instances of VOs in time). It is assumed that each frame of an input video sequence is segmented into a number of arbitrarily shape image regions (the VOs), and that the shape, motion and texture information of the VOPs belonging to the same VO is encoded and transmitted into separate VOLs corresponding to specific temporal or spatial resolutions (which allows later to separately decode each VOP and leads to the required flexible manipulation of the video sequence).

The three types of frames processed by such a coding structure are the following : the I-VOPs, the P-VOPs and the B-VOPs. An I-VOP is an intra coded VOP, the coding operation using information only from itself (it is the VOP that costs the most bits). A P-VOP is a predictive coded VOP, and the coding operation then uses a motion compensated prediction from a past reference VOP which can be either an I-VOP or another P-VOP (contrary to an I-VOP, only the difference between the current motion-compensated P-VOP and its reference is coded : thus, a P-VOP usually costs fewer bits than an I-VOP).

A B-VOP is a VOP that is coded using a motion-compensated prediction from past and future reference VOPs (I or P-VOPs). Said predictions are based on so-called forward and

backward motion estimations respectively. A B-VOP cannot be a reference VOP and, like the P-VOP, only the difference between the current motion compensated B-VOP and its reference VOP is coded.

Unfortunately, using said B-VOP prediction (also called interpolated or bi-directional mode) is not always a gain in term of compression. If the compression can sometimes be improved by a factor of about 20 %, it can also in other cases be decreased by a drastic factor.

## SUMMARY OF THE INVENTION

It is then an object of the invention to propose an encoding method using this B-VOP prediction only when it is efficient.

To this end, the invention relates to an encoding method such as defined in the introductory part of the description, said encoding method moreover including a coding step of each VOP and, before said coding step, a motion estimation step performed between the current VOP and the previous one, said motion estimation step itself comprising a decision process based on the following sub-steps of :

- carrying out a motion estimation between a VOP number N (VOP N) and the previous one (VOP N-1) ;
- on the basis of said motion estimation, computing a so-called coherence factor, provided for quantifying the sequence motion ;
- on the basis of a comparison of said coherence factor with a predetermined threshold, taking a final decision on the type of the current VOP, said current VOP being a B-VOP or not according to the value of said coherence factor with respect to said threshold.

## BRIEF DESCRIPTION OF DRAWINGS

The present invention will now be described, by way of example, with reference to the accompanying drawings in which Fig.1 illustrates the main steps of the encoding method according to the invention.

## DETAILED DESCRIPTION OF THE INVENTION

An MPEG-4 encoder comprises several functional blocks, among which one or several memories, for outputting the VOPs in the transmitting order required by the standard (for example, if the input sequence is I B B P B B P..., the output order will be I P B B P B B....), and a motion estimator, for receiving the current VOP and the previous one (or reference VOP) and taking the decision of which kind of prediction will be implemented for the current VOP : no prediction for an I-VOP, forward prediction for a P-VOP, bi-directional prediction for a B-VOP.

Within said motion estimator, the following decision is implemented, according to the invention. First, the current VOP (number : N) is captured. Then a motion estimation is carried out between the VOP N and the previous VOP (number : N-1), and a factor named "coherence factor" is computed, in order to qualify the sequence motion, and compared to a

threshold. According to the result of the comparison, the VOP N is allowed, or not, to be a B-VOP. The final decision concerning the prediction mode is then taken, and the coding step of the current VOP (= I-VOP, or P-VOP, or B-VOP) can take place.

The coherence factor proposed for the comparison test may be for instance (without limiting the scope of the invention) expressed as the ratio of the sum of absolute differences (SAD) between motion vectors of a macroblock (estimated in 16 x 16 pixels mode or 8 x 8 pixels mode) and its predecessor in the same VOP with the similar sum for the previous VOP (it is here recalled that for a macroblock of size $k \times k$, the expression of the SAD is :

$$SAD = \sum_{i=0}^{k \times k} |A(i) - B(i)|$$

where B(i) and A(i) respectively designate the current macroblock considered and the macroblock in the reference VOP which matches the most in a search zone defined in said reference VOP).

CLAIM :

1. A method for encoding a sequence of video data according to a standard in which several types of data are identified, said data consisting of so-called video Object Planes (VOPs) that are either intra coded VOPs (I-VOPs), coded using information only from themselves, or predictive coded VOPs (P-VOPs), coded using a motion compensated prediction from a past reference VOP, or bidirectionally predicted VOPs (B-VOPs), coded using a motion-compensated prediction from past and future reference VOPs, said encoding method including a coding step of each VOP and, before said coding step, a motion estimation step performed between the current VOP and the previous one, said motion estimation step itself comprising a decision process based on the following sub-steps of :

- carrying out a motion estimation between a VOP number N (VOP N) and the previous one (VOP N-1) ;

- on the basis of said motion estimation, computing a so-called coherence factor, provided for quantifying the sequence motion ;

- on the basis of a comparison of said coherence factor with a predetermined threshold, taking a final decision on the type of the current VOP, said current VOP being a B-VOP or not according to the value of said coherence factor with respect to said threshold..

Abstract

The present invention proposes a dynamic allocation of B-frames, according to which, for each input frame, a pre-analysis stage first performs preliminary forward motion estimation between current and previous frames, leading to current motion field. Both current and previous motion fields are then used to evaluate a coherence factor, which compares the sums of local differences within current and previous motion field. If the coherence factor is greater than an empirically determined threshold, a decision is taken to process current frame as a P-frame, if not, it is considered as a B-frame, while it is considered as a B-frame if said factor is equal to or lower than the threshold.

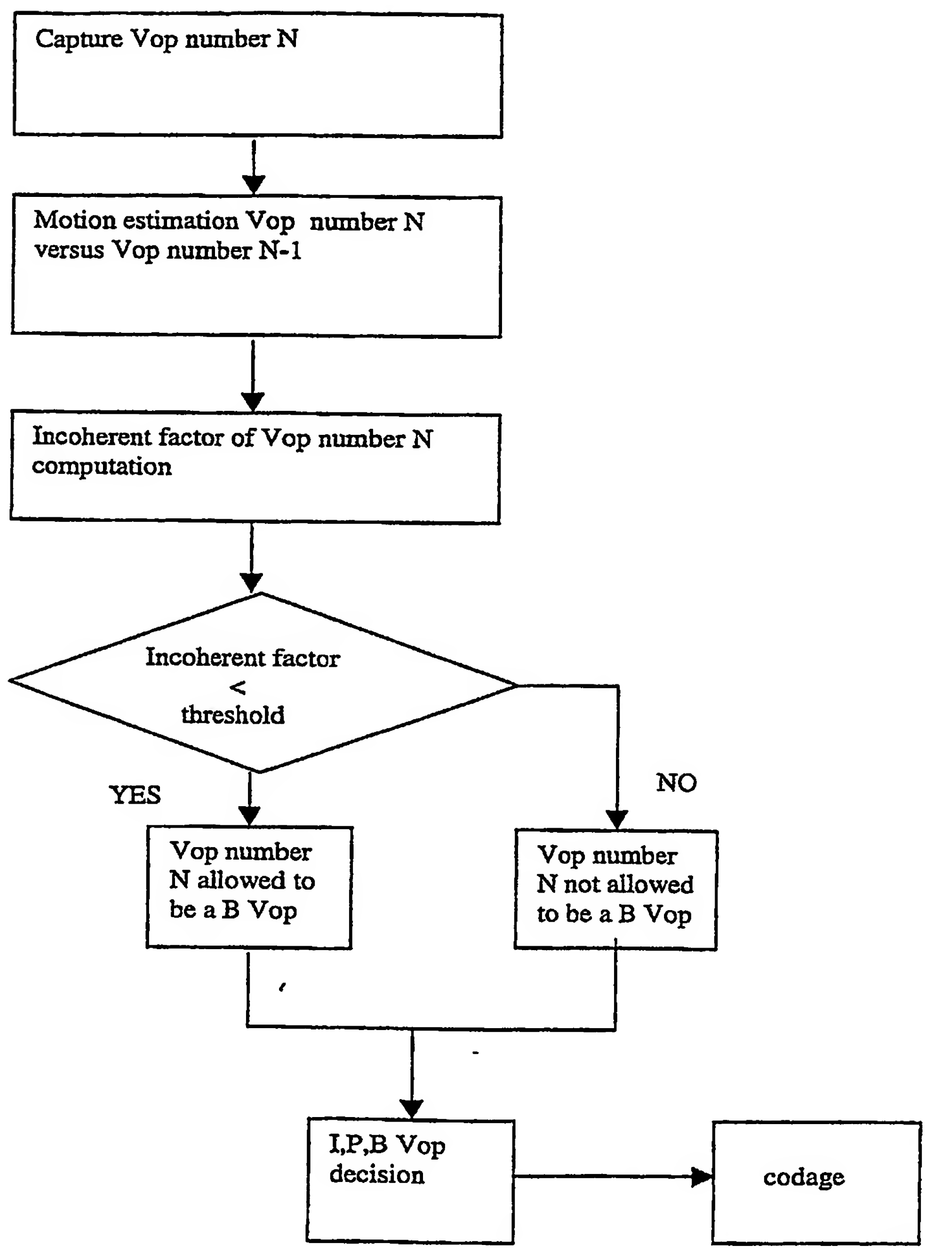
Capture Vop number N
Motion estimation Vop number N versus Vop number N-1
Incoherent factor of Vop number N computation
Incoherent factor
<
threshold
YES
NO
Vop number N allowed to be a B Vop
Vop number N not allowed to be a B Vop
I,P,B Vop decision
codage

FIG. 1